УДК 004.89:004.93

Ф.Г. Ващук, Ю.А. Василенко, І.Ф. Повхан, Л.С. Повхан
Закарпатський державний університет, м. Ужгород, Україна
zakdu@ukrpost.edu.ua

Проблема оцінки складності логічних дерев розпізнавання та загальний метод їх оптимізації

Дана робота є першою в циклі трьох статей, присвячених проблемі оцінки складності логічних дерев класифікації та розробці універсально підходу їх оптимізації. Проаналізований зв'язок логічних функцій та логічних дерев розпізнавання, на основі якого запропоновано досить простий спосіб мінімізації логічних дерев. Важливими перевагами даного способу мінімізації дерев є те, що з ним відносно просто працювати при великій кількості аргументів.

Вступ

Як відомо, більшість синтезованих систем розпізнавання (СР) у вигляді дерев класифікації можна записати або в ДНФ, або в КНФ формі. Так дерево розпізнавання, яке являє собою певне правило класифікації, можна представити за допомогою відповідної логічної функції $y = f(x_1, x_2, ..., x_n)$. Отже важливою проблемою при побудові СР такого типу буде проблема синтезу логічної функції, яка еквівалентна даному дереву розпізнавання [1], [2].

Так кількість усіх бінарних дерев, які фактично еквівалентні деякій логічній функції $f(x_1, x_2, ..., x_n)$, дорівнює $n!$ – кількості перестановок, серед яких знаходиться і саме складне. Під самим складним деревом будемо розуміти дерево, яке містить максимальну кількість різних міток, тобто якщо маємо дерева D і D^* та позначимо кількість міток на цих деревах через $S(D)$ і $S(D^*)$, зафіксувавши при цьому деяке n, то дерево D^* буде максимальним (найскладнішим) тоді і тільки тоді, коли для нього буде виконуватися рівність $S_n(D^*) = \max S_n(D)$.

Головною метою даної роботи буде мінімізація фіксованого класу самих складних дерев, причому нас буде цікавити знаходження не мінімальної форми, а найбільш ефективний спосіб мінімізації (перестановки ярусів), який дає найбільше число скорочених міток, іншими словами, спосіб, який значно зменшує складність початкового логічного дерева [1], [3].

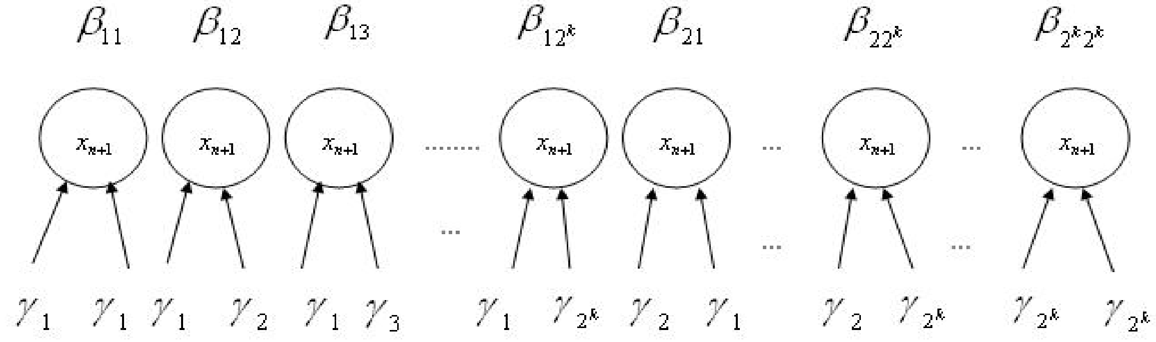
Задача полягає в проведенні такої мінімізації нижчевказаним способом до кінця та знаходженні ефекту мінімізації, тобто відношення $\frac{D_0}{D_S}$, де D_0 – найбільш складне дерево, з якого починається процес мінімізації, а D_S – дерево, отримане на S-му (останньому) етапі мінімізації.

Загальна схема логічної дерева

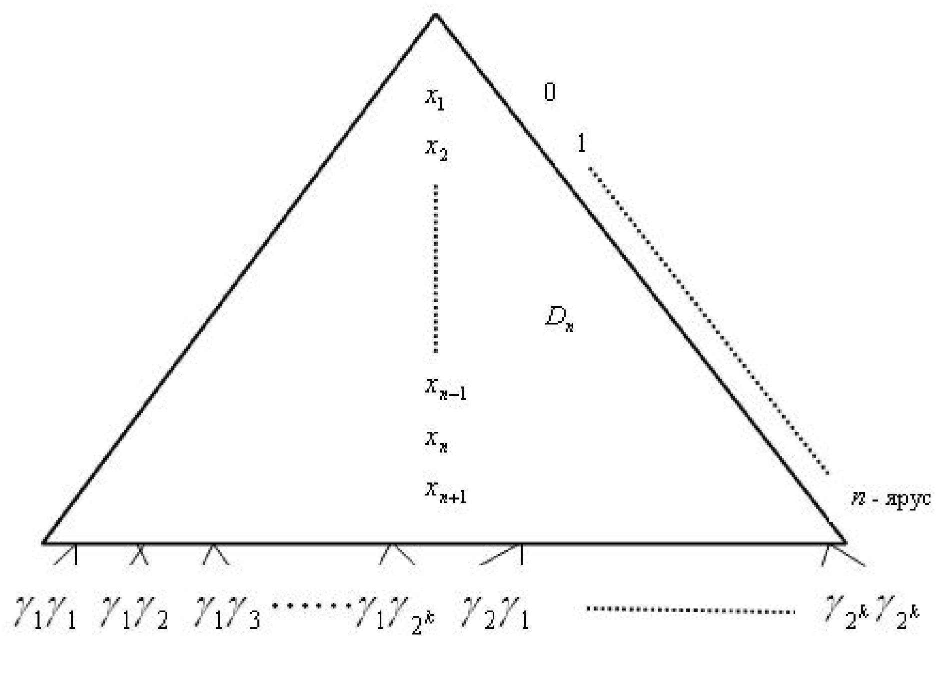
Перш за все побудуємо найбільш складне дерево.

Зауваження 1

Домовимося для зручності відлік ярусів починати з нульового. Нехай є довільне дерево D_n (блок розмірності $(n+1)$ змінних) з впорядкуванням $\beta_{ij}(i, j = 1, 2, ..., 2^k)$ на останньому ярусі, причому $\beta_{ij} \neq \beta_{ji}$, (рис. 1а, рис 1б).



а)



б)

Рисунок 1(а, б) – Дерево D_n розмірності $(n+1)$ змінних з впорядкуванням β_{ij}

Причому $\gamma_1, \gamma_2, ..., \gamma_{2^k}$ – мітки, які також побудовані якимось способом. Але вони є функціями, залежними від змінних $x_{i+1}, x_{i+2}, ..., x_{i+m}$, які знаходяться на нижніх ярусах (причому це твердження справедливе для довільної мітки дерева). В такому

випадку добудоване дерево $2 \cdot 2^{2^m} = 2^{2^m+1}$ буде залежати вже від N змінних, де $2^0, 2^1, 2^2, ..., 2^N$ – кількість вершин, а $2^{2^N}, 2^{2^{N-1}}, 2^{2^{N-2}}, ..., 2^{2^{N-N}}$ – кількість функцій (міток) відповідно на $0, 1, 2, ..., N$ ярусах.

Слід відзначити наступний дуже важливий момент – зі збільшенням номера ярусу кількість вершин збільшується, а кількість функцій зменшується, тому природно зробити припущення, що настане такий момент, коли на якомусь *i*-му ярусі кількість вершин (2^i) стане рівною кількості функцій $(2^{2^{N-i}})$, які можна розставити в цих вершинах, тобто:

$$2^i = 2^{2^{N-i}}. \tag{1}$$

Критичний ярус логічного дерева

Введемо поняття критичного ярусу.

Визначення 1

Ярус i довільного дерева D_n називається критичним, якщо всі його мітки різні.

Легко бачити, що ярус довільного дерева буде критичним тоді і тільки тоді, коли для нього буде виконуватися умова (1). З усього вищесказаного випливає наступна лема.

Лема 1

Якщо на *i*-му ярусі всі функції (мітки) різні між собою, то на 1, 2, ..., $(i-1)$-му ярусі функції також будуть різними.

Зауваження 2

В подальшому критичний ярус будемо позначати через K, тоді для критичного ярусу маємо наступне співвідношення:

$$2^K = 2^{2^{N-K}}. \tag{2}$$

Таким чином, для того, щоб добудоване дерево D_n було найбільш складним, необхідно, щоб воно мало максимальне число різних міток, що, в свою чергу, рівносильне тому, щоб дерево D_n було найбільш складним, а це можливо в тому і лише в тому випадку, (при даній структурі дерева), якщо всі γ_i, які знаходяться на *n*-му ярусі, будуть різними, тобто якщо *n*-й ярус буде критичним, так-як тоді (відповідно лемі 1) всі мітки, які стоять вище критичного ярусу, будуть також різними. Але вибране нами впорядкування $\beta_{ij} (\beta_{ij} \neq \beta_{ji})$ на останньому ярусі таке, що забезпечує різні γ_i на *n*-му ярусі, тому *n*-й ярус можна прийняти за критичний. Для справедливості даного твердження достатньо показати, що для ярусу n умова (2) виконується, а для $(n+1)$-го яруса – вже ні. Знайдемо клас логічних дерев, які задовольняють рівність (2), позначивши $N-K$ через m.

$$N - K = m. \tag{3}$$

де N, K, m – цілі числа. Підставивши (3) у (2), отримуємо:

$$2^K = 2^{2^m}. \tag{4}$$

З (4) випливає (5).

$$K = 2^m. \tag{5}$$

Підставивши значення K з (5) у (3) будемо мати $N - 2^m = m$, а отже

$$N = 2^m + m. \qquad (6)$$

Формула (6) при $m = 1, 2, ..., m_i$ з врахуванням формули (5) дає нам клас найбільш складних дерев.

Таблиця 1 – Клас найбільш складних дерев

m	$K = 2^m$	$N = 2^m + m$
1	2	3
2	4	6
3	8	11
.	.	.
.	.	.
m_i	2^{m_i}	$2^{m_i} + m_i$

Номер критичного ярусу знайдемо з рівняння (7).

$$2^x = 2^{2^{N-x}}, \qquad (7)$$

де $x = 2^{N-x}$, оскільки $N = 2^m + m$, то $x = 2^{2^m + m - x}$ і остаточно:

$$x = 2^m = K. \qquad (8)$$

Зробивши заміну в (2) *K* його значенням з (8), отримуємо рівність (2) в такому виді:

$$2^{2^m} = 2^{2^m}. \qquad (9)$$

Підрахувавши кількість вершин і функцій (міток) на $(n+1)$ ярусі, відповідно твердженню $n = k$ маємо $N = 2^m + m = k + m = n + m$ та

$$N = n = m. \qquad (10)$$

Тоді кількість вершин дорівнює кількості розгалужень

$$2 \cdot 2^{2^m} = 2^{2^m + 1}, \qquad (11)$$

а кількість функцій (міток) буде

$$2^{2^{N-(n+1)}} = 2^{2^{(N-n)-1}} = 2^{2^{m-1}}. \qquad (12)$$

Порівнявши (11) та (12), бачимо, що для $(n+1)$-го ярусу рівність (2, 9) буде порушуватися.

$$2^{2^m + 1} > 2^{2^{m-1}}. \qquad (13)$$

(кількість вершин) (кількість функцій)

Отже *n*-й ярус буде критичним, тобто всі γ_i, які знаходяться на $n[K]$ ярусі, різні. Таким чином, побудоване дерево D_n буде самим складним. Разом з $D_n^{'}$ воно буде належати класу самих складних дерев, в якому N визначається за формулою (6).

Зауваження 3

Кількість міток на $K+1$ $(n+1)$ ярусі дорівнює $2^{2^{m-1}}$ (12), оскільки для побу-

дови впорядкування β_{ij} ми використали 2^k міток: $\gamma_1, \gamma_2, ..., \gamma_{2^k}$, – то зрозуміло, що

$$k = 2^{m-1}. \qquad (14)$$

Оцінимо складність дерева D_n. Під складністю довільного дерева будемо розуміти кількість всіх різних міток дерева. Розрахунок будемо проводити наступним чином: знаючи, що на кожному ярусі $(0, 1, 2, ..., n[K])$ кількість вершин відповідно дорівнює $2^0, 2^1, 2^2, ..., 2^K$, а так як мітки, які знаходяться при даних вершинах, різні, то і загальна кількість всіх міток буде дорівнювати сумі вершин на кожному ярусі, тобто

$$2^0 + 2^1 + 2^2 + ... + 2^K.$$

За формулою геометричної прогресії $S = \dfrac{b_n q - b_1}{q-1}$ для $q > 1$ (в нашому випадку: $b_1 = 2^0 = 1, b_n = 2^K, q = 2$), маємо наступну оцінку складності дерева D_n:

$$2^0 + 2^1 + 2^2 + ... + 2^K = \frac{2^K \cdot 2 - 1}{2-1} = 2^{K+1} - 1 = 2^{2^m+1} - 1,$$

$$D_n = 2^{2^m+1} - 1. \qquad (15)$$

Дослідимо зміни складності дерева D_n, вважаючи його за початкове дерево $(D_n = D_0)$. Скористаємося теоремою (без доведення) про перестановку змінних (ярусів), як приклад взявши функцію двох змінних [3-5].

Теорема 1

$x_1(x_2(\gamma_1\gamma_2); x_2(\gamma_3\gamma_4)) = x_2(x_1(\gamma_1\gamma_3); x_1(\gamma_2\gamma_4))$, тобто при перестановці ярусів x_1 та x_2 мітки γ_2 та γ_3 міняються місцями (рис. 2).

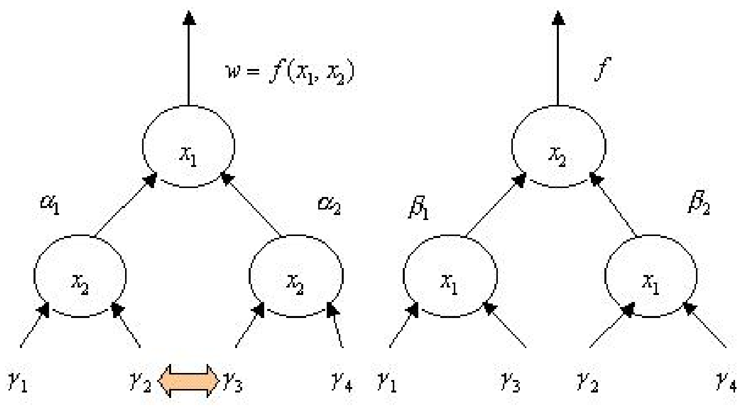


Рисунок 2 – Схема перестановки ярусів при перестановці змінних

Так як задача полягає в знаходженні деякого способу мінімізації, який давав би зменшення складності, то природно підібрати таку структуру розташування міток на деякому i-му ярусі, щоб при мінімізації (перестановці ярусів) отримати найбільшу кількість скорочень міток. Найбільш ефективною структурою є наступне розташування міток [6], [7].

$$\gamma_1\alpha\gamma_2\alpha......\alpha\gamma_{2^k}\alpha. \qquad (*)$$

Висновки

Тобто, якщо маємо дерево зі вказаною структурою (*), то пробивши визначене число кроків мінімізації, ми можемо розкласти його на два піддерева (підблоки), які залежать від меншого числа змінних. Зважаючи на це загальна схема методу мінімізації фіксованого класу логічних дерев базується на теоремі 1 та буде більш детально проаналізована в наступній частині роботі.

Література

1. Василенко Ю.А. Алгоритмическое конструирование распознающих систем на основе метода разветвленного выбора признаков (метод РВП) / Ю.А. Василенко // Тез. докл. Третьей Всесоюзной конференции «Математические методы в распознавании образов». – Львов, 1987. – С. 52-53.
2. Мінімізація логічних деревоподібних структур в задачах розпізнавання образів / І.Ф. Повхан, Ю.А. Василенко, Е.Ю. Василенко [та ін.] // European Journal of Enterprise Technologies. – 2004. – № 3(9). – С. 12-16.
3. Construction and optimization of recongnizing systems / Yu.A. Vasilenko, E.Yu. Vasilenko, A. Kuhayivsky [та ін.] // Інформаційні технології і системи. – 1999. – Т. 2, № 1. – С. 122-125.
4. Повхан І.Ф. Концептуальна основа систем розпізнавання образів на основі метода розгалуженого вибору ознак / І.Ф. Повхан, Ю.А. Василенко, Е.Ю. Василенко // European Journal of Enterprise Technologies. – 2004. – № 7(1). – С. 13-15.
5. Повхан І.Ф. Метод розгалуженого вибору ознак в математичному конструюванні багаторівневих систем розпізнавання образів / І.Ф. Повхан, Василенко Ю.А., Василенко Е.Ю. // Штучний інтелект. – 2003. – № 7. – С. 246-249.
6. Витенько И.В. Схемы, алгоритмы и многообразия / Витенько И.В. – Ужгород : Ужгород. ун-т, 1970. – 76 с.
7. Вітенько І.В. Математична логіка / Вітенько І.В. – Ужгород : Ужгород. ун-т, 1971. – 210 с.

Ф.Г. Ващук, Ю.А. Василенко, И.Ф. Повхан, Л.С. Повхан

Проблема оценки сложности логических деревьев распознавания и общий метод их оптимизации

Данная работа первая в цикле трех статей, посвященных проблеме оценки сложности логических деревьев классификации и разработке эффективного подхода их оптимизации. Проанализирована связь логических функций и логических деревьев распознавания, на основании которых предложен довольно простой способ минимизации логических деревьев. Важными преимуществами данного метода минимизации деревьев есть то, что с ним относительно просто работать при большом количестве аргументов.

F.G. Vashchuk, J.A. Vasilenko, I.F. Povhan, L.S. Povhan

Problem of Estimation of Complexity of Logic Trees of Recognition and the General Method of their Optimization

The given work is the first in a cycle of three articles devoted to the problem of estimation of complexity of logic trees of classification and elaboration of universal approach of their optimization. Connection of logic functions and logic trees of recognition is analysed and on basis of this connection enough simple way of minimization of logic trees is offered. Important advantages of the given method of minimization of trees is that it is rather easy to work with it at a considerable quantity of arguments.

Стаття надійшла до редакції 15.11.2010.